DOCUMENT RESUME

ED 076 728 HE 004 061

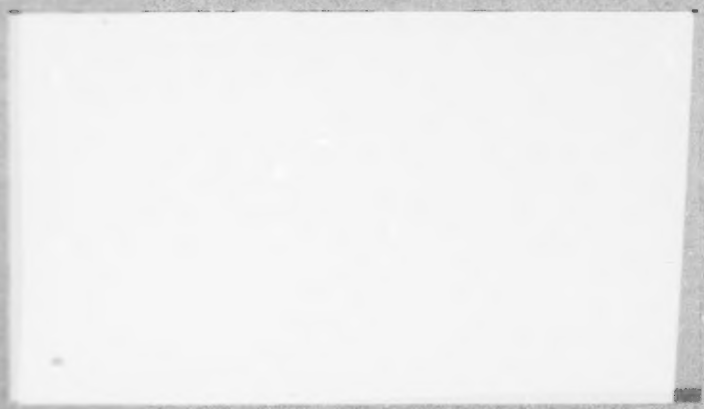
AUTHOR	Blackburn, Robert T.; Lingenfelter, Paul E.
TITLE	Assessing Quality in Doctoral Programs: Criteria and Correlates of Excellence.
INSTITUTION	Michigan Univ., Ann Arbor. Center for the Study of Higher Education.
SPONS AGENCY	Carnegie Foundation for the Advancement of Teaching, New York, N.Y.; State Univ. of New York, Albany. Office of the Regents.
PUB DATE	Feb 73
NOTE	54p.
AVAILABLE FROM	Center for the Study of Higher Education, The University of Michigan, Ann Arbor, Michigan 48104 ($1.00)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Doctoral Degrees; *Doctoral Programs; *Educational Quality; *Evaluation Criteria; *Evaluation Techniques; Graduate Study; *Higher Education; Program Evaluation; Research; Research Projects

ABSTRACT

This study, commissioned by the New York Board of Regents, concerns the evaluation of doctoral programs. Part one discusses critical problems in the assessment of excellence. Part two reviews a catalog of criteria and assessment techniques. Emphasis is placed on reputational studies, objective indicators of excellence, efficiency as an index of quality, client satisfaction ratings, intra-institutional evaluation, and evaluation of alumni. Part three indicates necessary conditions for adequate doctoral programs, and part four suggests a model for the evaluation of doctoral education. An extensive bibliography and appendices of related research material are included. (MJM)

FILMED FROM BEST AVAILABLE COPY

ED 078728



CENTER FOR THE STUDY OF HIGHER EDUCATION

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

HE 004 061

SCHOOL OF EDUCATION/THE UNIVERSITY OF MICHIGAN/ANN ARBOR

ED 078728

Price $1.00

ASSESSING QUALITY IN DOCTORAL PROGRAMS: CRITERIA AND CORRELATES OF EXCELLENCE

By

Robert T. Blackburn

and

Paul E. Lingenfelter

Center for the Study of Higher Education
The University of Michigan
Ann Arbor, Michigan 48104

February, 1973

We are appreciative of constructive critiques from Meredith Gonyea, Ralph Lewis, James L. Miller, Marvin Peterson, and Jack Waggett. The study reflects their concerns. This inquiry was commissioned by the New York State Board of Regents Commission on Doctoral Education who were supported in their study by the Carnegie Foundation. We wish to express our thanks to the Foundation whose support made this work possible. We assume full responsibility for what appears within this monograph.

TABLE OF CONTENTS

	Page
Preface by Wilbur J. Cohen, Dean, School of Education	i.
Prologue	ii.
I. CRITICAL PROBLEMS IN THE ASSESSMENT OF EXCELLENCE	1
II. A CATALOG OF CRITERIA AND ASSESSMENT TECHNIQUES	3
A. Reputational Studies	
B. Some Objective Indicators of Excellence	
C. Efficiency as an Index of Quality	
D. Client Satisfaction Ratings	
E. Intra-institutional Evaluation	
F. Evaluation of Alumni	
III. NECESSARY CONDITIONS FOR ADEQUATE DOCTORAL PROGRAMS	19
IV. A MODEL FOR THE EVALUATION OF DOCTORAL EDUCATION	20
Appendices	23
A: A Critique of Reputational Studies.	
B: Formulas for Aggregating Institutional Scores from the ACE Ratings.	
C: Formulas to Weigh Books and Articles in Scholarly Production Indices.	
D: The National Science Board Method of Calculating Correlates of Quality.	
E: The Hay Efficiency Index	
F: Intra-institutional Evaluation Programs	
References	37

PREFACE

At the University of Michigan School of Education we place great value upon research and service, in addition to instruction. This monograph is a result of some of the research and service which we hope will contribute to the betterment of American education.

I believe Professor Robert Blackburn and doctoral student Paul Lingenfelter have reached some significant conclusions in this study. Both men are associated with the School of Education Center for the Study of Higher Education, which has made a number of important researches into the nature of higher education.

The study described in this monograph concerns the evaluation of doctoral programs and was commissioned for the New York Board of Regents, who have been examining the rate of production of doctoral students in that state. I believe that aspects of the study presented in this monograph can have application to the examination of doctoral programs elsewhere.

This study is one in a long series conducted by the School of Education over its more than fifty-year history. We look forward to continuing to contribute in a variety of areas of research.

Wilbur J. Cohen

Dean

PROLOGUE

A desire for highest quality seemingly needs no defense, irrespective of the product or craft involved. Yet, the consideration of "quality" in any setting seems to provoke controversy. Whenever evaluative judgments are made and a rank order is established, someone is offended. Hence, the evaluator, sometimes correctly, sometimes in error, is likely to be charged with elitism, bias, racism, or a host of other unsavory traits.

Perhaps a part of the controversy regarding "quality" traces to ambiguities within the concept itself. After all, as Cartter (1966:4) notes, "in an operational sense, quality is someone's subjective assessment, for there is no way of objectively measuring what is in essence an attribute of value."

The following analysis avoids the use of "quality" in the general sense of the word. Although "quality" for any individual may be more than the sum of its parts, communication is facilitated and controversy is mitigated when "quality" is defined in terms of its component parts. Therefore, we have focused upon "excellence," defined or operationalized as dimensions of "quality." The constellation of "excellent" traits necessary for "quality" can be selected by the evaluator. However, he may find it more useful to let excellence stand on its own merits without striving to specify the more elusive characteristic, "quality." If the objectives of a specific program are defined, an assessment of its quality principally considers the degree to which those objectives are attained -- the degree to which excellence obtains along specific dimensions.

Two related matters. One is that there is agreement that quality and excellence exist, that some programs and products have more of it than do others.

Second, quality and excellence do matter. Further, they matter in important ways.

Studies show unequivocally that faculty behavior is associated with institutional quality (e.g., Wilson, 1942; Parsons and Platt, 1968), that institutions differ on scales of excellence (e.g., Berelson, 1960), and on numerous other characteristics, as the analysis which follows shows.

Also, despite the controversy which sometimes accompanies assessment, systematic evaluation of doctoral programs is becoming increasingly important and useful. Widely accepted projections of Ph.D. supply and demand (Brode, 1971; Cartter, 1971; National Science Foundation, 1971; Wolfe and Kidd, 1971) predict surplus doctorates in almost every field. Although these projections have drawn criticism (Letters, Science, 1971; Moses, 1972) (economic forecasting is never foolproof, especially given the possibility of behavioral market adjustments such as decreased applications for doctoral study), the magnitude of the predicted surplus is too great to be dismissed lightly. Financial and market pressures are likely to force cutbacks and reallocations of resources in doctoral education. Careful regular assessment of excellence is necessary to insure that whatever actions are taken relative to doctoral programs be guided by a clear perception of their strengths, weaknesses, and social contribution. Limited resources should be invested where they can produce the greatest return.

Of course, assessment is necessary for positive reasons as well. The expansion of knowledge continues unabated, and the rates of change in society and in the modes of transmitting and utilizing knowledge are accelerating without pause. Sound research and teaching at the doctoral level is critical both to develop needed technological and social innovations and to cultivate deeper understanding of the human situation. Given the rate of change in the environment, regular reevaluation of the effectiveness of graduate education is necessary to forestall obsolescence and irrelevance.

Finally, quality assessment possesses an internal virtue of no small consequence, one which by itself justifies the incorporation of program evaluation as an operational procedure as regular, say, as the annual audit. The assessment process qua process has salutary consequences. New ideas emerge; better practices are introduced, concerned self-analysis questions long standing assumptions, protects against dysfunctionalism, and, vitally important, generates a climate for healthy growth and development.

For these and other concerns of government and industry and education, excellence in graduate schools is a sine qua non. Regular assessment, no matter how agonizing at times, is fundamental, especially in graduate education. How to assess quality, then, is no idle concern.

I. CRITICAL PROBLEMS IN THE ASSESSMENT OF EXCELLENCE

Two problems confront the assessment of excellence in doctoral education: 1) Ascertaining the appropriate criteria for excellence; and 2) quantifying the criteria so as to permit comparisons among specific programs.

The selection of criteria is by no means an easy task. It is plagued both by political and conceptual difficulties. For obvious reasons, individuals and organizations favor criteria which focus on their own strengths. Since institutions compete for students, faculty, and funds, a commonly accepted definition of excellence has political complexities quite aside from conceptual ones. Notwithstanding the political problems, even a dispassionate,disinterested observer has serious difficulty selecting a set of non-contradictory criteria. His efforts meet a barrage of conflicting evidence and competing arguments.

In order to establish criteria of excellence, program evaluators first must decide what they value. From their values they then fashion objectives, which in turn, establish the criteria of excellence. Criteria vary with the objectives sought. For example, evaluators whose primary objective is the production of new knowledge establish criteria which define excellence for Ph.D. programs in terms of faculty and student scholarly output. They will rate a program with a distinguished faculty which produces much valuable research and a few outstanding, research-orientated Ph.D.'s much higher than evaluators who hold the training of college teachers as the primary objective of

Ph.D. programs. Appropriate criteria for the second evaluators might emphasize the numbers of qualified college teachers produced by a program rather than the quality of faculty and student research.

Of course, the objectives of program evaluators are rarely unidimensional. Furthermore, different objectives usually are neither mutually exclusive nor inclusive. For example, it may not be possible to maximize both the generation of outstanding scholarly work and the production of competent college teachers. However, it may be possible and desirable to maintain an optimal mix by balancing competing values in single institutions or by seeking different objectives in different programs (Warren, 1967). Ideally the criteria established for evaluation of doctoral programs will be formed by a conscious review of all relevant objectives and a conscious weighting of those objectives on the basis of an hierarchy of values. The established value hierarchy, and consequently the criteria for evaluating Ph.D. programs, will vary with the perceived needs of the institution, state or nation. Certainly it is possible for doctoral programs emphasizing somewhat different objectives to attain excellence and to receive due rewards.

After the basic task of selecting explicit criteria has been completed, the second critical problem in evaluation is encountered. What is a valid measure of relative degrees of excellence on a stated criterion? For example, what is a valid measure of a faculty's scholarly abilities? Is it peer evaluations? Research grants obtained? Number of publications?

What is a valid measure of departmental "effectiveness?" Is it alumni salaries? Placement of graduates? Ph.D.'s per faculty? The list is long.

While most of the techniques necessary for evaluating doctoral programs have already been developed or can be developed relatively easily, problems of measurement cannot be dismissed lightly. However, where measurement difficulties exist, the weakness of one technique may be offset by the strengths of another. Hence, by utilizing several measures of effectiveness a relatively comprehensive and valid evaluation can occur.

Each of the three elements of the evaluative process -- objectives, criteria, and assessment indices -- have unique problems. But the most critical prerequisites of successful evaluation are that: 1) Each element of the process be clearly defined; and 2) the relationships between objectives, criteria, and indices be logical and explicit. If these conditions are not met, the evaluative process creates unnecessary confusion and controversy.

II. A CATALOG OF CRITERIA AND ASSESSMENT TECHNIQUES

A. Reputational Studies

The most widely known, heralded, and criticized evaluations of doctoral programs have been the reputational studies. Hughes (1925) conducted the first of these in the 20's. Similar studies have followed his lead. (Hughes, 1934; Keniston, 1959; Cartter, 1966; Roose and Anderson, 1971).[1] The two most recent studies (1966 and 1971), sponsored

---

[1] Hughes (1946) published a third study of doctoral education, but it was

not a reputational survey.

---

by the American Council on Education, have been methodologically the most sophisticated and the most widely publicized. Higgins (1968: 8-55) effectively summarizes the history of graduate school assessment in his unpublished rating of doctoral programs in education.

Peer evaluation characterizes all the repuational studies. A panel of scholars rates the quality of an institution's faculty and/or graduate program in a given discipline. By combining individual panelist's responses, an aggregated rating of the graduate program is calculated.

Although the criterion of excellence which dominates reputational studies is the scholarly ability of a faculty, the technique of peer rating can be used to obtain measures of other criteria. For example, in addition to scholarly ability the ACE studies used "program effectiveness" as a criterion. Other criteria, such as "supportiveness of graduate students," "effectiveness of teaching," or "student quality" could be used. However, since widely scattered outside evaluators are less likely to be aware of conditions pertinent to such criteria than they are of a faculty's scholarly output, reputational assessments or other criteria can not be defined as easily as can assessments of scholarly ability.

In addition, reputational studies suffer from a lack of normative standards. Two matters confound peer ratings of excellence. First, almost by definition only a few institutions can receive "outstanding" ratings. Hence, what results is more a rank order of excellence rather than a position on a normative scale. Second, standards of excellence change over time. For example, a university rated average today may have improved dramatically over the recent past while its position relative to other

institutions does not reflect its improvement

Reputational evaluations have been widely criticized (Lewis, 1968; Council of Graduate Schools, 1970; Bess, 1972; Elton and Rose, 1972). However, for what they primarily assess, scholarly competence, no better measure is available. Nevertheless, criticisms of reputational studies are not dismissed out of hand. A more complete discussion of these criticisms and the characteristics of the various reputational studies appear in Appendix A.

Although the ACE studies rated only departments (contending that this is the largest unit in a university which can be accurately evaluated (Cartter, 1966: 106)), others have aggregated the ACE scores and published composite ratings of entire institutions (Ewell, 1966; Magoun, 1966). Magoun (1966) argues that such composite ratings are useful indices of effective central administration. (Several formulas for aggregating departmental ratings appear in Appendix B.)

B. Some Objective Indicators of Excellence

Several observers of graduate education have selected one or more objective characteristics as benchmarks of excellence. Some objective characteristics, labeled "correlates of quality," have been established by examining various programs rated in the ACE studies and distilling a cluster of objective traits associated with high quality ratings. Other evaluators have selected indicators on an *a priori* basis, defining quality in terms of certain characteristics and proceeding to develop an index of those characteristics. Calvert, *et al* (1971) combine several objective indicators to construct a general index of quality. The latter of these approaches is discussed first.

1) Scholarly Productivity: When a critical objective is the production of new knowledge, scholarly output is an important criterion of excellence. Indices of scholarly productivity have been developed for both the evaluation of individuals and departments. The typical index assigns weights to books, articles, reports, patents, etc., and a productivity score is calculated by aggregating the quantity of credits in each category. Examples of productivity indices can be found in the work of Crane (1965), Pelz and Andrews (1966), Stallings and Singhall (1971), and others. (Appendix C summarizes various schemes for weighting different types of output.)

Several researchers have found a relationship between the most prolific departments and those receiving top rankings in reputational studies (Berelson, 1960: 127; Cartter, 1966: 80,88,101; Crane, 1965: 703-705). Productivity, however, changes with time. An appreciable body of research examines faculty productivity, teaching effectiveness, receptivity to new ideas and adaptability as they relate to age, rank, and tenure. Since a number of variables affect performance, the development of a productive faculty mix must take into account the relevant factors. The research is summarized in Blackburn (1972).

An indirect, but possibly more valid index of scholarly productivity is the citation index used by Clark (1957), Bayer and Folger (1966), Cole and Cole (1967), Creager (1967), and Myers (1970). Since scholarly contribution in the strictest sense of the word requires that other scholars utilize a piece of work, this index calculates the number of citations of the work of a research or departmental faculty which appear in subsequent scholarly publications. Citations of work

that is more than ten or fifteen years old are given extra weight, on the assumption that durability is an indicator of unusual quality. The citation method is particularly useful in the natural sciences since an index of citations, the Science Citation Index, is maintained by the Institute for Scientific Information (Garfield and Scher, 1965). In other academic disciplines, particularly in the humanities, the citation technique may be less useful because citation of published work is not as common, and no such index of citations now exists. Margolis (1967) and Smith and Fieldler (1971) discuss the strengths and weaknesses of the citation index in greater detail.

2) Degrees, Awards, and Other Faculty Traits: Some observers have used other faculty characteristics as indicators of excellence. Bowker (1965) examined the number of awards won by a faculty as an index of its quality. Perkins and Snell (1962: 114-118) examined the percentage of history faculty with the Ph.D., years of experience, the percentage of instructors teaching graduate students, the percentage of faculty publishing in the past five years, the percentage publishing on subjects other than their Ph.D. dissertation topic, and the percentage of foreign specialists who have traveled abroad.[2] Indicators

---

[2] The Perkins and Snell study was not primarily a rating of individual departments but rather a description of graduate study in history as a whole. However, the characteristics listed above are implicitly utilized as indicators of quality.

---

such as these can play a useful role in evaluating graduate departments.

However, they are not as useful as measures more closely related to the actual productive work of faculty, such as scholarly writing or the training of Ph.D.'s.

3) Student Quality: Another technique for evaluating doctoral programs has been to measure the quality of students enrolled in a program. In one sense this is a reputational measure, because faculties tend to select the best possible students and good students are attracted to programs with a reputation for quality. In another sense, however, well-qualified students are an essential element of an excellent program. Thus student quality can stand in its own right as a criterion of excellence.

Perkins and Snell (1962: 38-39) compared history graduate students to students in other fields on the basis of undergraduate GPA, I.Q. scores, and Graduate Record Examination Scores. These indices can easily be used to compare students in different doctoral programs in the same discipline. Bowker (1965) utilized the distribution of Woodrow Wilson fellows among graduate programs as an index of quality. Cartter (1965: 107-112) used the distribution of student-attached fellowships (Woodrow Wilson, NSF, and others) as a cross-validation of the ACE study.

Although the number of prizes obtained by the students of a program is an interesting and useful datum, all graduate students do not compete for awards. In addition, errors can occur in the selection process of fellowships. For example, the Ph.D. success rate for Woodrow Wilson Fellows is not impressive (Mooney, 1968: 52,53).

standardized test scores and past academic performance seem to be more useful indicators of student quality. Astin's (1965) study of undergraduate college selection can be a useful model.

However, even standard tests will require modification to mitigate their cultural bias when used with minority group members. Several graduate programs have developed effective but unorthodox instruments to measure academic ability in minorities who do not fare well on standardized tests (Gunne and Leslie, 1972). Particularly since equalized opportunity is a desirable social goal, measures of student ability which do not allow for cultural bias are inadequate indicators of quality in institutions which are enrolling large numbers of minority group members. Also, grade point averages are related to grading practices which are not uniform across the country. As colleges adjust high school GPA's on the basis of experience, so do graduate departments weigh differentially undergraduate grading as practiced by different colleges.

As yet no objective evidence exists to demonstrate relationships between the existence and/or percentage of faculty who are women and/or minority group members and the attraction and success rate of students from client groups heretofore discriminated against. However, limited experience and role theory suggest positive connections. Assuming the expansion of opportunities for minorities in doctoral programs is socially desirable, the absence or presence of minority group faculty as mentor-models must be considered in the assessment of a program.

4) Physical Facilities: An obvious facility required for doctoral

study is an adequate library. Several writers consider library strength an important criterion of excellence (Perkins and Snell, 1962: 137: Jordan, 1963; Cartter, 1965: 114-115). Jordan (1963: 374) found that highly rated institutions have larger libraries and spend more per student on librarian salaries.

More than a superficial review of library resources is necessary, however, if the library is to be used as an index of quality. A specific number of volumes and a budget of reasonable size are necessary for excellence in many disciplines, but they may not be sufficient to assure excellence. A massive library can be inadequately supplied in a given discipline. It may not be up-to-date, or it may be administered in a way which discourages its utilization. Only a careful review can determine whether a library is adequate for doctoral work.

No systematic studies of other physical facilities for doctoral education (laboratories, office space, computer capabilities, seminar rooms, etc.) exist, possibly because these are rarely designed solely for graduate instruction. However, evaluation of such facilities is appropriate in the assessment of new or existing programs. Since the facilities required differ among disciplines, specialists must be utilized for the evaluation of physical facilities.

5) Correlates of Reputational Quality: The National Science Board (1972) conducted a major study of graduate education which, among other things, attempted to identify the factors closely associated with quality graduate programs as ranked in the ACE studies. The following factors are found to be correlates of reputational quality:

1) Magnitude of the doctoral program (number of degrees awarded).
2) Amount of federal funding for academic research and development.
3) Non-federal current fund income for educational and general purposes.
4) Baccalaureate origins of graduate fellowship recipients (NSF fellowships).
5) Baccalaureate origins of doctorates.
6) Freshman admissions selectivity.
7) Selection of institutions by recipients of graduate fellowships (NSF fellowships).
8) Postdoctoral students in science and engineering.
9) Doctoral awards per faculty member.
10) Doctoral awards per graduate student.
11) Ratio of doctorates to baccalaureate degrees.
12) Compensation of full professors.
13) The proportion of full professors on a faculty.
14) Higher graduate student/faculty ratios.
15) Departmental size of seven faculty members or more, (this finding is not a strict correlate calculated from median scores.) (National Science Board, 1969: 49-108) (c.f. Elton and Rogers, 1971).

(Appendix D. describes the method used to calculate correlates in this study.)

Two independent kinds of data are related to the NSB correlates. First, the notion of mutually strong supporting disciplines (e.g., mathematics for physics) is shown by Berelson (1966: 124). High intercorrelations exist in the ratings of departments in an area of study (e.g., r = .76 in social sciences). Second, while no one debates the need for a critical mass of faculty, arguments continue on whether there is an optimum faculty (and hence student) size. Wispé (1969) and Elton and Rose (1972) report positive correlations

between ACE ratings and size. They imply size may be the "cause" of the rating inasmuch as bigger is more likely to insure identification and hence insure reputation. Gallant and Prothero (1972) find no overall correlation between size and Roose and Anderson ratings in five selected disciplines. In fact, one analysis they conduct suggests an optimum size; that is, quality drops when a certain size is exceeded. Hagstrom (1971) finds positive correlations with department size and the 1966 ACE ratings (.57). In fact, he obtains positive correlations with a number of the variables already mentioned -- journal article output (.67), research opportunities (.45), a citation index (.69), number of post doctoral fellowships (.63), as well as selectivity in undergraduate and Ph.D-granting institutions, awards, offices, and morale.

Given the impressive list of correlates above, a temptation exists to assert that high value on these characteristics are necessary and sufficient conditions of quality doctoral education. The authors of the National Science Board (1969:108) report are careful to note that "the simulation of the values of a group of factors will not ensure a graduate program of high quality." That is to say, the mere existence of the correlates is not a sufficient condition of quality. Although high values on some of the correlates may be necessary for high quality (e.g., numbers 2, 12, and 15 are likely candidates), a causal connection is not proved by correlation analysis. However, calculating correlates of quality is useful for understanding the conditions which usually accompany good doctoral programs. Close examination of these programs can reveal exactly what contribution the correlated factors make to high quality.

C. Efficiency as an Index of Quality

Some definitions of excellence in doctoral education omit efficiency in any

form as a criterion. Others focus on efficiency and exclude many other criteria. Given the scarce resources currently available to higher education, efficiency is an indispensible criterion of excellence.

Despite conventional wisdom, the most "efficient" doctoral programs are not necessarily low quality diploma mills, nor are the least "efficient" (in degree production terms) necessarily top quality programs that slowly and deliberately nuture their doctoral students until they attain the highest standards of scholarship. The National Science Board evidence suggest that institutions with high reputational ratings produce more Ph.D.'s per graduate faculty member and per student enrollment than lower rated institutions (National Science Board, 1969: 76,79). The greater efficiency of these programs may be due to economies of scale, support, or a host of other factors. In any event, the evidence indicates that "quality" and "efficiency" are not mutually exclusive. Just the opposite, in fact, is indicated.

The Council of Graduate Schools has conducted a wide-ranging study of the costs and benefits of graduate education (Powel and Lamson, 1972).[3]

---

[3] A very useful annotated bibliography has been compiled in conjunction with this study.

Their task was hampered by 1) inadequate information regarding the actual costs of graduate education and 2) exceedingly complex conceptual problems in establishing a valid measure of the social benefits of graduate education. The ideal quantitative analysis of the costs and benefits of graduate education (and many other social welfare functions) is beyond our capacity at this time, and the necessary techniques are

likely to elude us in the foreseeable future. For these reasons and the fact that a duplication of the CGS study is unnecessary, we make no effort to discuss efficiency on the broad social level. Rather, we concentrate upon methods to assess the relative efficiency of similar Ph.D. producing educational programs.

In the early sixties, when gloomy prognostications of Ph.D. shortages were widely accepted, several studies examined the factors which contributed to attrition and to the inordinately long time required by some students to complete their work. Using enrollment data and questionnaires from deans, faculty, current students, successful Ph.D.'s, and drop-outs, these studies found that the humanities, followed by the social sciences, consistently required longer periods of study for the doctorate and experienced higher rates of attrition (Berelson, 1960: 157; Wilson, 1965: 69; Tucker, 1969: 64-67). Breneman has developed a theory of external market relationships and internal incentives to explain the persistence of this pattern among disciplines (Breneman, 1972).

Other efforts to measure efficiency have focused upon the doctoral production rate of a faculty. Blackburn and Trowbridge (1972) examined the workload implications of dissertation chairmanships and found wide variations in the productivity of individuals and departments. The National Science Board (1969) used the ratio of Ph.D.'s to graduate faculty and to student enrollments as measures of efficiency. Hay (1970) has developed an index of Ph.D. productivity which weighs the distribution of rank in a faculty (a faculty with more full professors is expected to produce more Ph.D.'s), considers student enrollment (including

the length of study in terms before the Ph.D.), and calculates productivity ratios which may be compared across departments. (Appendix E provides the details of this technique.)

The ideal index of efficiency in Ph.D. production probably has not been devised. However, a good one can be developed from the efforts which have already been made. It must include: 1) Enrollment data for students from the time of entry until the termination of their study (with or without a degree); 2) tabulations of individual and departmental activity relative to dissertation committees; 3) tabulation of undergraduate work loads; and 4) tabulation of all instructional activities (seminars, directed readings, etc.) relative to doctoral education. Computer-assisted analysis of such data is relatively simple; gathering it in standarized form from widely scattered institutions and departments is more difficult. In the event that detailed data is available, the next best alternative is to utilize gross measures after the manner of the National Science Board study.

D. Client Satisfaction Ratings

An obvious means of evaluating an enterprise of any kind is to ask those it serves for their opinions. The "consumers" of doctoral education are students and the employers of graduates. Both of these groups have been utilized in evaluative studies.

Alciatore and Eckert (1968: 11-53) asked a group of Ph.D. recipients to evaluate their training at the University of Minnesota. Their opinion of their preparation for specific tasks (teaching, research, etc.) was solicited as well as their general satisfaction with their career and

the doctoral program that launched it. A similar survey was conducted at Florida State University (1957). Such surveys of student opinion are most valuable when they include both current and past, successful and unsuccessful students so that a wide range of opinion is included in the study.

Student opinion surveys have been part of several major studies of graduate education. Berelson (1960: 275) surveyed 3,843 recent doctoral recipients in order to gain information about their experiences and opinions about possible reforms. (Heiss, 1970: 274-303) conducted extensive surveys of student opinion at ten major graduate schools and devoted much of her study to the analysis of their suggestions and complaints. Tucker's (1964) study of attrition, Gottleib's (1961) study of graduate student socialization, and several other research efforts have used graduate student or alumni opinions as explicit evaluative instruments (Hughes, 1959; Mechanic, 1962; Darlington, 1970; Toombs, 1971). Harvey (1972) reviews most of the work on graduate students.[4]

---

[4] Hall (1971) provides an excellent annotated bibliography on graduate education in general which includes a number of student studies.

The other major "clients" of doctoral programs are the employers of graduates. McGrath (1961: 11-16) used a panel of liberal arts college presidents to assess the performance of graduate schools in the production of college professors. About 75% of the presidents felt

that new Ph.D.'s came to their institutions without knowledge and skills important to their job. Similar surveys could obtain evaluations from all employers of Ph.D.'s -- government, business, and all levels of education -- and relate them to specific doctoral programs.

E. Intra-Institutional Evaluation

During the past few years many graduate schools have instituted formal evaluations of academic departments. (Appendix F includes information about several of these.) Most review efforts utilize a committee of administrators, outside faculty, and/or students as the primary review body. In many cases their function is primarily to provide feedback to a department in order to encourage improvement. However, at Minnesota and Illinois major evaluation efforts have been undertaken in a climate of fiscal emergency.

Review committees generally have access to rudimentary statistical information and the indicators of excellence mentioned above, but the unique instrument for evaluation at their disposal is wide-ranging observations and interviews with faculty and students within a program. Less tangible indicators, e.g., student and faculty morale or teaching effectiveness, can be assessed these ways. Carefully corroborated impressions gathered from such a review can anticipate problems in seemingly strong programs or detect promise in a developing department. Although such techniques require funding, they can be used selectively to evaluate programs which are neither unusually good nor bad on standarized indicators.

Review committee evaluations can be generated by a state agency

as well as within an institution, or a state board can merely ask for the results of internal self-studies as part of its total evaluation program. The latter approach may be ill-advised, however, since the possibility of outside sanctions can discourage candor and thoroughness in institutional self-studies. Assured confidentiality enhances candid self-evaluation, reduces defensiveness, and provides a powerful impetus for improvement. Since intra-institutional evaluation obtains these advantages, an external assessment program must take care not to vitiate its effectiveness.

F. Evaluation of Alumni

The new Ph.D. is the basic output of a doctoral program. Some means of assessing the quality of these persons is essential to an evaluation program. Fieldler and Biglan (1969) examined the placement of new Ph.D.'s in their first job as an indicator of quality. This approach tests a program's reputation in the job market. After employers are rated as to desirability, the placement survey is used to assess the competitive position of a program vis à vis similar programs. The technique also can be used to assess the appropriateness of academic preparation to the work actually being performed by new Ph.D.'s.

In addition to placement surveys and the surveys of employer and recipient satisfaction mentioned previously, Ph.D. alumni may be evaluated on the basis of their scholarly productivity using the techniques listed above. Crane (1965: 704, 705) and Berelson (1960: 126) found that the most productive scholars tend to be alumni of prestigious programs. Another forthright measure of Ph.D. quality

is content analysis of dissertations. This has not been done systematically, but expert outside readers can provide an unbiased estimate of quality by reviewing student dissertations or even dissertation abstracts. An example of a content analysis study is Persell (1972).

III. NECESSARY CONDITIONS FOR ADEQUATE DOCTORAL PROGRAMS

Professional opinion often argues that certain conditions must be met in order to have an adequate doctoral program. These conditions most often include coverage of certain specialties within a discipline (e.g., both organic and inorganic in chemistry) as well as the presence of good supportive departments in related disciplines (e.g., mathematics for physics). But because necessary conditions of quality are difficult to establish empirically, competent professionals occasionally disagree on the precise definition of what is absolutely essential.

This issue is not as critical when evaluating an existing program. Then past productivity can be examined. However, essential conditions of excellence must be considered when evaluating a new program proposal. With trained advice to establish minimum levels for specific disciplines, the following are essential:

1) Are adequate resources available to teach doctoral students?
2) Is the departmental faculty large enough to provide necessary stimulation from a variety of perspectives?
3) Are essential sub-fields covered by the faculty?
4) Are the library and physical resources adequate?
5) Do the scholarly achievements of the faculty indicate that they can handle advanced instruction in the proposed areas?
6) Is instruction in supporting disciplines available and adequate?
7) Is there an available supply of good students?

The guidelines for advanced studies published by the North Central Association (1970) contain a discussion of these questions.

IV. A MODEL FOR THE EVALUATION OF DOCTORAL EDUCATION

The above catalogue of evaluative techniques is useful only as individual indicators of excellence find their appropriate place in an integrated conceptualization of doctoral education. Figure 1 graphically displays such a conceptualization and the place of individual indicators of excellence in the broader system.

[Insert Figure 1]

Although input, process, and output are interrelated, it is important to maintain distinctions between these categories when evaluating doctoral programs. A program with all of the requisite input values can be weak if its faculty is torn by dissension or demoralizes students with ambiguous requirements and debilitating criticism. Or a department can conceivably rank well on input and process measures and yet produce output of little social value. For example, an evaluation of a highly productive program in Slavic Languages might indicate the desirability of a cutback simply because there are few jobs for Ph.D.'s in Slavic Languages.

At the state level, decision-makers must decide what it is they want from doctoral education. After this critical decision is reached, an existing program can be evaluated by examining its output and assessing the efficiency of its processes. When a proposed program is

Figure 1

VARIABLES RELEVANT TO EXCELLENCE IN DOCTORAL PROGRAMS

	INPUT VARIABLES	PROCESS VARIABLES	OUTPUT VARIABLES
	1) Student ability 2) Faculty ability 3) Financial support 4) Physical facilities 5) Supporting disciplines	1) Morale a. Faculty b. Student 2) Efficiency a. Faculty b. Student 3) Pedagogical methods a. Curriculum content b. Instructional methods 4) Research procedures	1) Faculty research 2) Student research 3) Ph.D. alumni productivity
(Indicators of Excellence)	1) Student ability (standarized tests, GPA, fellowships held) 2) Faculty ability (scholarly reputation, past productivity of research and Ph.D.'s, doctoral training.) 3) Financial support (salary levels, "soft" money, etc.) 4) Physical facilities (library volumes, library budget, lab facilities, etc.) 5) Supporting disciplines (must exist and meet quality indicators)	1) Morale (surveys, interviews) 2) Efficiency (degrees awarded per faculty member, attrition, completion time, workload studies) 3) Pedagogical methods (internal reviews, outside visitors.) 4) Research procedures (internal reviews, outside visitors.)	1) Faculty research (publicatio count, citation indices, reputational studies) 2) Student research (publica-tion count, citation indice dissertation reviews) 3) Ph.D. alumni productivity (placement in first job, employer satisfaction, recipient opinion polls, scholarly productivity)

examined for accreditation or an existing program fails to meet the desired objectives, a careful assessment of both inputs and processes is necessary for diagnosis. Whether the prescription is curative or euthanasic will depend upon the cost of vitalizing the program and some assessment of the social need or demand for its services. Neither the infusion of additional resources nor euthanasia should be prescribed until a careful assessment is made of all aspects of a doctoral program.

APPENDIX A: A Critique of Reputational Studies

Although a systematic criticism of reputational studies has not been published, attacks on this technique generally criticize the impartiality of the evaluating panel, the subjectivity of reputation, the competence of the panel, or the validity of the criteria used in reputational studies. Although these criticisms do not apply with equal force to the most recent reputational studies, each criticism and a defense of reputational methodology will be considered below.

First, the question of panel impartiality. It is a well-known fact tha a handful of doctoral programs are alma mater to the majority of all Ph.D.'s. Although this pattern is changing, in 1934, Hughes' top fifteen institutions accounted for 59% of all doctorates and in 1957 Keniston's top fifteen produced 43% (Berelson, 1960: 97). At least one reviewer equates size with quality (Eells, 1957), but the high correlation between size and high reputational ratings may be explained in part by rater bias (see Wispé, 1969; Elton and Rose, 1972). It is understandable that a professor would favor the department in which he is teaching or from which he received his degree in a reputational survey. Consequently, departments with a large faculty and alumni group will receive higher ratings, so the argument goes.

Other sources of bias may come from the procedure used to select evaluators. Keniston used department chairmen, and Hughes used his faculty at Miami University (1925) and the secretary of national scholarly associations (1934) to select his panel. Department chairmen may not be representative of professional opinion in a discipline, and relying upon a single expert to select a panel risks contamination from the biases of that individual.

The ACE studies reduced the possibility of rater bias by systematically selecting a balanced panel of evaluators. Graduate deans at all institutions included in the evaluation were asked to select both junior and senior faculty to participate on the panel. The large number of deans participating and the balanced distribution of rank in the panel largely eliminates the possibility of systematic bias through panel selection procedures.

The possibility of bias through current or past affiliations is less easily mitigated, but in a large study it is relatively unimportant. Cartter's careful internal analysis of his data revealed that, as expected, raters tend to be biased toward their alma mater and current employer, but such biases tend to have negligible effect. The high split-half reliability of Cartter's ratings suggest that systematic bias in the evaluating panel has been prevented by careful techniques of panel selection and by utilizing a large number of respondents.

The second major argument against reputational studies is the intrinsic subjectivity of reputation. Quoting Dr. Johnson, one respondent to the Cartter survey observed that "a compendium of gossip is still gossip" (Cartter, 1966:8). In a similar vein, some argue that "subjective" reputational studies are inferior to evaluations based upon "objective" traits

such as faculty publications, library facilities, etc. (Lewis, 1968). As Cartter noted, however, "objective" criteria are based upon subjective notions of quality one step removed (Cartter, 1966: 4). Any effort at evaluation requires a degree of trust in the subjective wisdom of the evaluations, whether they assess a program directly or merely establish other indicators as criteria of quality. (In defense of reputational studies it should be noted that high ratings tend to be correlated with "objective" measures (Cartter, 1966).

Third, some contend that a valid assessment of a doctoral program can only be made from first hand exposure to the program (Council of Graduate Schools, 1970: 129-130). Since only a fraction of a survey panel is likely to have had direct exposure to more than a small fraction of the total population of programs, their assessments, perforce, are of dubious validity. This argument holds more weight when considering ratings of all aspects of a program than when considering ratings of a single attribute such as the scholarly ability of a faculty. Particularly in this case, faculty in a discipline have firsthand exposure to the scholarly work of their peers all over the country. The competence of scholars to evaluate the scholarly work of others is almost self-evident, even though their competence to assess other aspects of doctoral programs may be questioned if they lack firsthand information.

Fourth, and finally, the criteria of quality used in reputational studies may be challenged. The ACE studies used the "quality of the graduate faculty (defined in terms of scholarly achievements) in your field" and the overall "effectiveness[5] of the doctoral program" as criteria for the

---

[5]"Effectiveness" is defined as "the accessibility of faculty and their scholarly competence, curricula, educational and research facilities, the quality of graduate students, and other (relevant) factors."

---

ratings (Cartter, 1966: 127). Since the first criterion is the most widely cited in reviews of the reputational studies, many argue that scholarly ability alone is an insufficient condition for "quality." Other factors such as teaching effectiveness and efficiency are important as well. The second criterion, as defined, encompasses all other relevant variables, but as mentioned above, the fact that most evaluators lack direct contact with most doctoral programs casts doubt upon the validity of ratings in this category.[6] Nevertheless, scholarly competence is a necessary, if not

---

[6]The high correlation between "scholarly achievements" and "effectiveness" ratings suggests that raters chose to emphasize the trait they knew best when evaluating "effectiveness."

---

sufficient, condition of excellence and the ability of reputational surveys to measure this attribute has not been sufficiently challenged.

One other factor related to criterion validity deserves mention. In a reputational study the criteria deemed important by the raters must be accepte

implicitly. In most cases where professional consensus obtains, a non-professional observer is, and should be, inclined to accept the judgment of the professional panel. But, if a discipline is undergoing fundamental changes in its methodology or theoretical frameworks, or if a discipline on the whole has failed to adapt itself to changing environmental needs, the values of faculty raters may be unreliable or incongruent with those of a non-professional observer.

In political science during the late fifties the controversy between professors interested in behavioral approaches and those concerned with normative structural analysis quite likely was reflected in the ratings. Although partisans of either perspective may have been counterbalanced by the other, interpretation of ratings in a divided discipline is risky without some direct indication of the criteria emphasized by individual raters. And, given the possibility that an entire discipline embraces dysfunctional values, a non-professional should have at least some knowledge of professional norms before accepting reputational ratings uncritically.

To summarize, reputational studies have been criticized on the following grounds: 1) Panel bias; 2) Subjectivity; 3) Panel competence; and 4) Criterion validity.

In defense of reputational studies we have suggested that: 1) Panel bias has been largely eliminated by the careful selection procedures of the ACE studies; 2) Subjectivity cannot be escaped in evaluation no matter what technique is used; 3) Professional peers are competent to evaluate scholarly work, the central criterion in reputational studies; and 4) Although not a sufficient condition of general excellence, scholarly ability is necessary for a good doctoral program.

Reputational studies cannot provide a comprehensive assessment of overall quality, and they deserve critical examination whenever used, but for ratings of the scholarly ability of a faculty they are a valuable and valid instrument of assessment.

APPENDIX B: Formulas for Aggregating Institutional Scores from the ACE Ratings

Magoun (1966) calculated institutional ratings in the broad divisions of study (humanities, social sciences, etc.) by adding the ACE scores given to departments in a single institution. For example, the rating of an institution's social science departments on the ACE scale of 0-5.0 would be added together to create a total score. Exact ratings are given by the ACE for scores in the 3.0 - 5.0 range, and Magoun scored departments placed in the "good" or "adequate plus" categories 2.0 and 1.0 respectively. Lower categories were scored .50. A divisional score or institutional score equals the total scores of its departments.

Ewell's (1966) scoring scheme assigned points to departments based upon its rank order and its general classification in the ACE study. Departments in the distinguished-strong groups were given 14 points for making that category plus their inverse rank in the category. In a field of 17 distinguished or strong departments the top ranking department would be assigned 31 points, the second ranking 30 points, and the 17th ranking department would receive 15 points. Departments rated "good" received 10 points, and those rated adequate received 5 points. Institutional scores are calculated by adding the scores assigned to individual departments.

The McCurdy methodology was devised to assist the AAU in ranking prospective member institutions. McCurdy examined only the traditional liberal arts and science departments and assigned points by the category of evaluation received in the 1969 ACE study. Both the quality of the graduate faculty and the effectiveness of the graduate program ratings were included. Table 1 outlines the McCurdy formula (Fawcett, 1971).

All of the above formulas, except McCurdy's, work to the advantage of large institutions with many specialized departments. An institution with 25 adequate or good departments could receive a higher score than an institution with six or seven departments all of which were rated good, strong, or distinguished. In an effort to correct for this bias, Wallis, (1971) devised a technique which compared two institutions only on the basis of departments they have in common. The advantage of this technique is that it controls for size; the disadvantage is that a separate score must be calculated for each institution in the comparison group. (E.g., Michigan has one score relative to Columbia, another relative to Yale, another relative to Berkeley, etc.) Wallis' formula is:

$$\frac{(\#\text{ of departments in which A is rated higher than B+ } \#\text{ of ties})}{(\#\text{ of departments in which B is rated higher than A+ } \#\text{ of ties})}$$

A simpler approach to controlling for size in institutional ratings is to add individual department scores (using the Magoun, Ewell, or McCurdy methods) and divide by the number of departments in an institution to obtain an average quality score.

The various schemes for aggregating departmental ratings generally do not produce radically different rank orderings in institutions. If standard errors of measurement are considered, whatever differences do obtain are likely to be insignificant.

Table 1: Revised McCurdy Points Methodology

1. Institutions listed in the Distinguished-Strong Category . . . 5 Points

2. Institution's Departments (Fields of Discipline) Rated by Quality of Graduate Faculty (According to 1969 A.C.E. Report)

   a. Distinguished Rating . . . . . 4 Points
   (The Top 5 schools, including ties, in each Department)

   b. Strong Rating . . . . . 3 Points
   (Residual of schools with points of 3.0 or above but below the Distinguished Rating assumed as Top 5)

   c. Good Rating . . . . . 2 Points
   (Between 2.0 & 2.9 in '69 A.C.E)

   d. Adequate Plus Rating . . . . . 1 Point
   (Between 1.0 & 1.9 in A.C.E.)

3. Institution's Departments (Fields of Discipline) Rated by effectiveness of Graduate Program (According to 1969 A.C.E. Report)

   a. Extremely Attractive . . . . . 3 Points
   (Between 2.0 & 3.0 in 1969 A.C.E.)

   b. Attractive . . . . . 2 Points
   (Between 1.5 & 1.9 in 1969 A.C.E.)

   c. Acceptable Plus . . . . . 1 Point
   (Between 0.8 & 1.4 in 1969 A.C.E.)

4. Total Points Summed for each institution and ranked accordingly

APPENDIX C: Formulas to Weigh Books and Articles in Scholarly Production Indices

A. Manis (1907)

1 book = 18 articles

B. Crane (1965)

Publications were classified as either major or minor and tabulated separately. A major publication was a book, a journal article of more than fifty pages, or four closely related journal articles. Minor publications were edited books, translations, jointly authored (more than three authors) books, research articles, and lab manuals. Short reports, book reviews, unpublished papers, and popular articles were not counted at all. If a professor had published a major work, his minor writings were not tabulated (Crane, 1965: 701-702).

C. Cartter (1966)

1 theoretical or research book	= 6 articles
1 textbook	= 3 articles
1 edited collection	= 2 articles
4 short communications	= 1 article
8 book reviews	= 1 article
16 book notes	= 1 article

D. Stallings and Singhal (1969)

PUBLICATION TYPE	WEIGHTED POINTS
Book (Author)	9
Book (Co-Author)	6
Dissertation	5
Book (Editor)	4
Article (Author)	3
Article (Co-Author)	2
Technical Report	3
Technical Report (Co-Author)	2
Book Review	2
Book Review (Co-Author)	1

E. Lazarsfeld and Thielens (1958: 402-407) suggest that the rank ordering of professors' productivity is affected only slightly by the formula used to weigh books and articles. Hence, a formula which assigns equal weight to all forms of publication should be completely adequate for most purposes of comparison. This is confirmed by Cartter's experimentation with various weighting schemes (Cartter, 1966: 81).

APPENDIX D: The National Science Board Method of Calculating Correlates of Quality

Standard correlation techniques were not used to distill the "correlates of quality" in the National Science Board report. Instead, the 1966 ACE rankings of graduate programs were aggregated into institutional scores, and institutions were assigned to one of seven quality classes, A,B,C,D,E,F, or G in descending order of excellence. Median scores on a number of factors were calculated for each quality class. If the median value of a factor descended uniformly (one or two exceptions were occasionally admitted) from class A to class G, the factor was deemed a quality correlate.

For example, the distribution of median ratios of graduate fellowship awards to number of applicants by quality was A, .360; B, .296; C, .273; D, .267; E, .248; F, .215; G, .197. Since these median values descend uniformly by quality class, the success ratios of fellowship applicants is considered a correlate of quality. If the distribution across quality classes had been markedly irregular or not uniformly descending, the characteristic in question would have been rejected as a quality correlate.

The advantage of this technique is that it simplifies the presentation of the evidence by displaying the correlates in graphic, easily understood form. Bar graphs or line graphs clearly show that as "quality" decreases so do the median values on the correlated factor. A disadvantage is that by departing from standard statistical methods potentially strengthening or damaging evidence is lost. Since no correlation coefficients are reported, no tests of significance are possible. Moreover, the shape of the distribution within a quality class can only be guessed (only median values are reported), and it is theoretically possible for actual correlation values to range from very high to very low levels while satisfying the test of correlation used in the study. It is possible that the National Science Board technique is superior to standard approaches because it removes the effects of grossly deviating cases within each quality class, but an outside observer would be better able to evaluate their evidence if the usual measures of correlation had been reported as well.

APPENDIX E: The Hay Efficiency Index

d = no. of doctorates awarded in year (year begins Sept. 1)

s = no. of distinct thesis directors involved in doctoral awards

n = no. of equiv. full professors in dept. (EFP) (Prof. = 1.0, Assoc. Prof. = 0.75, Asst. Prof. = 0.25)

i = no. of graduate students in Term I

I = average of i over 5 years

Y = no. of elapsed calendar years, baccalaureate to Ph.D.

E = no. of terms of graduate enrollment anywhere of successful students

D = average of d over 5 years

S = average of s over 5 years

N = average of n over 5 years

A = D/N = no. of doctorates awarded per year: per EFP

C = 100 S/N = % of EFP directing theses to completion in year

P = 100 D/I = % of graduate students receiving doctorate in year.

Note -- Using five year averages smooths the distribution curve and mitigates the impact of wide fluctuations in yearly productivity. A and C are measures of gross faculty productivity. P is a measure of a department's efficiency with regard to student resources.

APPENDIX F: Intra-institutional Evaluation Programs

Following are selected pages from a description of the internal evaluation program at the University of California, Berkeley. Extensive internal evaluation programs are being developed in many institutions. Minnesota and Illinois have launched assessment programs, descriptions of which should be available through the graduate dean's office at these institutions. Also, a major assessment program is being developed at Michigan. In several months detailed descriptions of this effort should be available.

University of California, Berkeley

PROCEDURES FOR REVIEW OF DOCTORAL PROGRAMS

The reviews of doctoral programs will be guided by two major commitments: that the improvement of graduate education can be fostered most effectively by collecting as much information as possible about each program under review; and that the same procedures will be applied to all programs reviewed. In accord with these principles, the following procedures will be used as a means for securing the types of information listed in the attached itemization.

In each case, a principal source of information will be the members of the department under review; other sources of information are designated in the itemization. It is anticipated that two forms of information will be solicited from department members: written and oral. Written information will include various formal departmental announcements and documents as well as less formal letters and reports. Information will be communicated orally during the course of meetings of the review panel with department members.

Within each department, three subgroups may be distinguished: the chairman and graduate advisers; the remainder of the faculty; and the graduate students. In order to promote candor in the communications from department members in these groups to the review panel, it is proposed that information be solicited separately from each group. Accordingly, in addition to whatever documents are submitted by members of the three groups, three separate meetings will be scheduled for attendance by the panel, one for each group. The first of these three meetings will be a dinner-discussion between the panel and the department chairman along with graduate advisers. At this time, the panel will detail its procedures for the chairman who can also use the occasion to present and discuss his own views of the doctoral program in his department. Daytime meetings will involve the remaining two groups: faculty and graduate students. A final meeting will also be scheduled to which members of all three groups will be invited. This meeting will permit the panel to present its report for discussion and comment by all components of the department prior to its presentation to the Graduate Council. Prior to preparing a final version of the report, comment on a preliminary version will also be invited from appropriate officers of other relevant units (e.g. Dean of the College).

In accord with this plan, each review will be initiated by letters from the Chairman of the Graduate Council and the Dean of the Graduate Division to the members of the three departmental groups (as well as to two other pertinent groups: Teaching Assistants and other non-Senate members of the instructional staff). These letters will invite the recipients to the appropriate meetings, outline the review procedures to be followed, and describe the kinds and forms of information the panel wishes to secure from members of each group. (Samples of these letters are enclosed).

The membership of the review panel will vary depending on the identity of the department under review. Every panel will include the Chairman of the Graduate Council and the Dean of the Graduate Division. The Council Chairman will appoint one or two other members of the Council to each panel, selecting persons from departments allied with that being reviewed. An Associate Dean of the Graduate Division may be appointed to the panel by the Dean in view of the relationship between the Associate Dean's discipline and that of the department to be reviewed. Finally, a graduate student may be designated to serve on the panel and to consult with graduate students in the department under review.

While the panel is securing information directly from departmental sources, the Graduate Division will take responsibility for securing and assembling all other information listed in the attached itemization. In addition, the Graduate Division will assist the panel in preparing its report to the Council.

University of California, Berkeley

GUIDELINES FOR EVALUATING DOCTORAL PROGRAMS

Type of Information	Source of Information
A. Description of Present Status and expected fate of disciplines in which work is offered	Chairman, Advisers, Faculty, Students
B. Program	
1. Original description	Graduate Division
2. Approved Revisions	Graduate Division
3. Unauthorized Revisions	Chairman and Advisers, Students
4. Informal Description	Chairman, Advisers, Students
5. Course Offerings	
a. Content, number, variety	Catalog
b. De facto scheduling	Schedule and Directory
C. Advising	
1. Availability-frequency	Chairman, advisers, other faculty present and former students
2. Quality	Same
D. Teaching	
1. Faculty load: formal and informal	Chairman, advisers, other faculty and schedule and directory
2. Effectiveness	Chairman, advisers, present and former students
E. Evaluation of Student Progress	
1. Methods of evaluating course work	Chairman, advisers, other faculty, students
2. Extra-course methods of evaluation: examinations -- character, objectives, frequency, sequence, failure rate; research requirements, thesis and dissertation requirements	Chairman, advisers, other faculty, students, Graduate Division

3. Course offerings

   a. content, number, variety

   b. De facto scheduling

Type		Source
F.	Admissions	
	1. Numbers admitted annually	Graduate Division
	2. Procedures	Chairman, advisers
	3. Criteria	Chairman, advisers
G.	Resources	
	1. Faculty	
	a. Number filled and vacant FTE	Chancellor
	b. Distribution by rank	Chancellor
	c. Turnover rate	Chancellor
	d. Retirement prospects	Chancellor
	e. Distribution by training institution	Chancellor
	2. Student Financial Support; TAs, RAs, Fellowships, Grants, etc.	Graduate Division
H.	Productivity	
	1. Quantity	Graduate Division: EDP
	a. Admission rate	
	b. Years to M.A.	
	c. Years to Doctorate	
	d. Rate of granting M.A. and doctorate	
	e. Aldrich-Hammel index	

2. Quality

a.	Sample of recent theses and dissertations	Library
b.	Letters from former students	Graduate Division
c.	Occupational history of former students	Chairman, faculty, Graduate Division, Alumni offices

I. Morale

1.	Faculty	Chairman, advisers, faculty
2.	Student	
a.	Degree and methods of participation in departmental decisions	Chairman, advisers, students, G.A. reps.
b.	General	Students

REFERENCES

Alciatore, Robert and Ruth E. Eckert

1968 "Minnesota Ph.D.'s Evaluate Their Training: A Study of the Relationship of Various Ph.D. Programs to Later Career Service and Satisfaction." University of Minnesota, Minneapolis, Minnesota: (October).

Astin, Alexander

1965 Who Goes Where to College? Chicago: Science Research Associates, Inc.

Bayer, A.E. and John K. Folger

1966 "Some Correlates of a Citation Measure of Productivity in Science." Sociology of Education 39 (Fall): 381-390.

Berelson, Bernard

1960 Graduate Education in the United States. New York: McGraw-Hill.

Bess, James L.

1971 "Ranking Academic Departments: Some Cautions and Recommendations." Journal of Higher Education 42 (December): 721-727.

Blackburn, Robert T. and Keith Trowbridge

1972 "Faculty Accountability and Faculty Workload: A Primary Cost Analysis of their Relationship as Revealed by Ph.D. Productivity." The University of Michigan. (Submitted for publication.)

Blackburn, Robert T.

1972 Tenure: Aspects of Job Security on the Changing Campus. Atlanta: Southern Regional Education Board (July).

Bowker, Albert

1965 "Quality and Quantity in Higher Education." Journal of the American Statistical Association 60 (March): 1-15.

Breneman, D.W.

1972 An Economic Theory of Ph.D. Production: The Case at Berkeley. Berkeley: Ford Foundation Program for Research in University Administration, Report P- 8 (December).

Brode, W.R.

1971 "Manpower in Science and Engineering, Based on a Satuation Model." Science 173 (July 16): 206-213.

Calvert, Jack G., James N. Pitts, Jr., and George H. Dorion

1971 Graduate School in the Sciences: Entrance, Survival, and Careers. New York: Wiley-Interscience.

Cartter, Allan M.

1966 An Assessment of Quality in Graduate Education. Washington, D.C.: American Council on Education.

Cartter, Allen M.

1971 "Scientific Manpower for 1970-1985." Science 172 (April): 132-140.

Clark, Kenneth E.

1956 America's Psychologists: A Survey of A Growing Profession. Washington, D.C.: American Psychological Association.

Cole, Stephen and Jonathan R. Cole

1967 "Scientific Output and Recognition: A Study in the Operation of the Reward System in Science." American Sociological Review 32 (June): 377-390.

Crane, Diana

1965 "Scientists at Major and Minor Universities: A Study of Productivity and Recognition." American Sociological Review 30 (October): 699-714.

Council of Graduate Schools in the United States

1970 Reassessment: Proceedings of the Tenth Annual Meeting. Miami Beach, Florida: (December).

Creager, J.A.

1967 The Use of Publication Citations in Educational Research. Washington, D.C.: American Council on Education.

Darlington, Paul E.

1970 "An Analysis of the Obstacles Perceived by Graduate Students as Delaying Their Progress Toward the Doctorate." Unpublished Ph.D. dissertation, The University of Michigan.

Eells, Walter Crosby

1957 "Leading American Graduate Schools." Association of American Colleges Bulletin 43 (December): 563-576.

Elton, Chas F. and Samuel A. Rogers

1971 "Physics Department Ratings: Another Evaluation." Science 174 (November 5): 565-568.

Elton, Charles F. and Harriett A. Rose

1972 "What are the Ratings Rating." American Psychologist 27 (March): 197-201.

Ewell, Raymond H.

1966 "A Quantified Summary of the American Councilor's Evaluation Report 'An Assessment of Quality in Graduate Education.'" State University of New York - Buffalo. (Mimeographed)

Fawcett, Novice G.

1971 Letter to Presidents of the Association of American Universities Members. Ohio State University (April, 12).

Fiedler, F. E. and Anthony Biglan

1969 The First Job After Graduation as A Measure of Departmental Performance. Technical Report 74. Urbana, Illinois: Group Effectiveness Research Laboratory (April).

Florida State University

1957 A Study of Florida State University Doctoral Graduates and Their Reactions to the Doctoral Program at the University. Tallahassee: Florida State University Press.

Gallant, Jonathan A. and John W. Prothero

1972 "Weight-Watching at the University: The Consequences of Growth." Science 175 (January 28): 381-388.

Garfield, E. and S. Scher

1965 Science Citation Index. Philadelphia: Institute for Science Information.

Gottlieb, David

1961 "Process of Socialization in American Graduate Schools." Social Forces 40 (December): 124-131.

Gunn, Manuel G. and Larry L. Leslie

1972 Exceptional Graduate Admissions to the Pennsylvania State University. Report 15. Center for the Study of Higher Education, Pennsylvania State University, University Park.

Hagstrom, W.O.

1971 "Inputs, Outputs, and the Prestige of University Science Departments," Sociology of Education 44 (Fall): 375-97.

Hall, Wayne C.

1971 An Annotated Bibliography on Graduate Education, 1950-71. Washington, D.C.: National Research Council.

Harvey, James

1972 The Student in Graduate School. Washington, D.C.: American Association for Higher Education.

Hay, George E.

1970 "Doctoral Productivity." Horace Rackham School of Graduate Studies, The University of Michigan, Ann Arbor, Michigan. (Mimeographed)

Heiss, Ann M.

1970 Challenges to Graduate Schools. San Francisco: Jossey-Bass.

Higgins, Arthur S.

1968 The Rating of Selected Fields of Doctoral Study in the Graduate Schools of Education: An Opinion Survey. Unpublished Ed.D. dissertation, Teachers College, Columbia University.

Hughes, E.C.

1959 "Stress and Strain in Professional Education." Harvard Educational Review 29 (Fall): 319-329.

Hughes, R.M.

1946 A Study of American Graduate Schools Conferring the Doctorate: 1937-38, 1941-42. Ames, Iowa: State College Press.

Hughes, R.M.

1934 "Report of the Committee on Graduate Institution." Educational Record 15 (April): 192-234.

Hughes, R.M.

1925 A Study of the Graduate Schools of America. Oxford, Ohio: Miami University.

Jordan, Robert T.

1963 "Library Characteristics of Colleges Ranking High in Academic Excellence." College and Research Libraries 24 (September): 369-376.

Keniston, Hayward

1959 Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania. Philadelphia: University of Pennsylvania Press.

Lazarsfeld, Paul F. and Wagner Thielens, Jr.

1958 The Academic Mind. Glencoe, Illinois: The Free Press.

Letters to the Editor

1971 Science 173 (July 2): 6,7.

Lewis, Lionel S.

1968 "On Subjective and Objective Ratings of Sociology Departments." The American Sociologist 3 (May): 129-131.

McGrath, Earl J.

1961 The Quantity and Quality of College Teachers. New York: Teachers College, Columbia University.

Magoun, H.W.

1966 "The Cartter Report on Quality in Graduate Education." Journal of Higher Education 37 (December): 481-492.

Manis, Jerome G.

1951 "Some Academic Influences upon Publication Productivity." Social Forces 29 (March): 267-272.

Margolis, J.

1967 "Citation Indexing and Evaluation of Scientific Papers." Science 155 (March 10): 1213-1219.

Mechanic, David

1962 Students Under Stress. Glencoe: Free Press.

Mooney, Joseph D.

1968 "Attrition Among Ph.D. Candidates: An Analysis of a Cohort of Recent Woodrow Wilson Fellows." The Journal of Human Resources 3 (Winter): 47-62.

Moses, Lincoln

1972 "The Response of Graduate Enrollment to Placement Opportunities." Science 177 (August): 494-497.

Myers, C.R.

1970 "Journal Citations and Scientific Eminence in Contemporary Psychology." American Psychologist 25 (November): 1041-1048.

National Science Foundation

1971 1969 and 1980, Science and Engineering Doctoral Supply and Utilization. Washington, D.C.: National Science Foundation.

National Science Board

1969 Graduate Education: Parameters for Public Policy, Washington, D.C.: National Science Foundation.

North Central Association

1970 Guidelines for Institutions Offering Advanced Degree Programs. Chicago: Commission on Institutions of Higher Education, North Central Association (December).

Parsons, Talcott and Gerald Platt

1968 The American Academic Profession: A Pilot Study. National Science Foundation, Grant GS513 (March).

Pelz, Donald C. and Frank M. Andrews

1966 Scientists in Organizations. New York: John Wiley.

Perkins, Dexter and John L. Snell

1962 The Education of Historians in the United States. New York: McGraw-Hill.

Persell, Caroline

1972 "The Quality of Research on Education: An Empirical Study of Researchers and Their Work." New York: Columbia University, Bureau of Applied Social Research.

Powel, John H., Jr. and Robert D. Lamson

1972 An Annotated Bibliography of Literature Relating to the Costs and Benefits of Graduate Education. Washington, D.C.: The Council of Graduate Schools (March).

Elements Related to the Determination of Costs and Benefits of Graduate Education. Washington, D.C.: The Council of Graduate Schools (March).

Roose, Kenneth D. and Charles J. Anderson

1971 A Rating of Graduate Programs. Washington, D.C.: American Council on Education.

Smith, Richard and Fred E. Fiedler

1971 "The Measurement of Scholarly Work: A Critical Review of the Literature." Educational Record 52 (Summer): 225-232.

Stallings, William M. and Sushila Singhal

1970 "Some Observations on the Relationships between Research Productivity and Student Evaluation of Courses and Teaching." American Sociologist 5: 141-143.

Toombs, William E.

1971 "Graduate Students and Sponsored Research: Selected Aspects of Socialization at the Dissertation Level." Unpublished Ph.D. dissertation, the University of Michigan.

Tucker, Allen

1964 Factors Related to Attrition Among Doctoral Students. East Lansing: Cooperative Research Project No. 1146, Michigan State University.

Wallis, W. Allen

1971 Letter to President Novice G. Fawcett of Ohio State University (January 28).

Warren, Donald

1967 "Interorganizational Field as A Focus for Analysis." Administrative Science Quarterly 12 (December): 224-237.

Wilson, Kenneth

1965 Of Time and the Doctorate. Atlanta: Southern Regional Education Board.

Wilson, Logan

1942 Academic Man. New York: Oxford University Press.

Wispé, Lauren G.

1969 "The Bigger the Better: Productivity, Size, and Turnover in A Sample of Psychology Departments." American Psychologist 24: 662-668.

Wolfe, Dael and Charles V. Kidd

1971 "The Future Market for Ph.D.'s." Science 173 (August): 784-793.